

# Workload Management: Stale files from MPI jobs filled /dev/shm, what now?

*Stale files from MPI jobs filled /dev/shm, what now?*

Sometimes when compute nodes stay up for long periods of time, /dev/shm gets filled with stale files. This can happen if MPI jobs abort in an unexpected way.

The stale files get cleaned up if the node is rebooted. A cleanup that avoids a reboot is also possible, simply by remounting /dev/shm, but this may affect MPI jobs using /dev/shm at that time.

A gentler way to deal with this is to have a script clean /dev/shm if needed. It can be run each time a job attempts to start by adding it as a custom prejob health check.

The following script deletes files under /dev/shm that don't belong to users that are running jobs on the node:

```
#!/bin/bash
```

```
SHMDIR=/dev/shm
```

```
# do not remove stale root files  
ignoretoken="-not -user root"
```

```
# get the users in the node via ps, as w/who don't work without login  
for user in $(ps -eo euid,euser --sort +euid --no-headers | awk '{if($1 > 1000) print $2;}' | uniq)
```

```
do  
    ignoretoken="${ignoretoken} -not -user $user"  
done
```

```
# clean up  
find $SHMDIR -mindepth 1 $ignoretoken -delete
```

The following steps add a prejob healthcheck via cmsH:

```
# cmsH  
% monitoring healthchecks  
% add clear_shm  
% set command /path/to/custom/script  
% commit  
% setup healthconf <category name>  
% add clear_shm  
% set checkinterval prejob  
% commit
```

Unique solution ID: #1161

Author: Frank Furter

Last update: 2013-12-11 18:03